A COMPARISON OF TEACHER RATINGS MADE BY STUDENTS WITH HIGH AND LOW DISCREPANCIES BETWEEN COURSE GRADE AND CUMULATIVE GRADE POINT AVERAGE

A Thesis

Presented to

the Faculty of the Department of Psychology

Appalachian State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Sylvia Southard Odom

May 1970

A COMPARISON OF TEACHER RATINGS MADE BY STUDENTS WITH HIGH AND LOW DISCREPANCIES BETWEEN COURSE GRADE AND CUMULATIVE GRADE POINT AVERAGE

by

Sylvia Southard Odom

May 18, 1970

Approved by:

Joyce G. Crouch
Chairman, Thesis Advisory Committee

Curtis Wilkins
Dean of the Graduate School

Mary Louise Powell
Major Professor

Walter T. Snipes
Minor Professor

ACKNOWLEDGMENT

It is with great appreciation for advice and guidance in the preparation of this study that I wish to acknowledge those who so graciously gave of their time.

I am deeply grateful to Dr. Joyce Crouch, Chairman of the Thesis Advisory Committee, for her continued advice, patience, and assistance in the preparation of this study; and to the other committee members, Dr. Walter Snipes and Mrs. Mary Louise Powell for their guidance and interest.

Gratitude is extended to the students who participated by rating their teachers and to the following faculty members who cooperated by allowing class time for rating and by supplying course grades: Dr. Basil Johnson, Dr. George Wesley, Mr. Irvin Kauffman, Dr. Willard Brigner, and Dr. Melvyn Gilley. Dr. Walter Snipes, Chairman of the Psychology Department, contributed by making arrangements with the Office of the Registrar for securing cumulative grade point averages.

In conclusion, I wish to thank my husband for his encouragement and patient understanding without which this endeavor would not have been possible.

TABLE OF CONTENTS

Chapter		PAGE
I.	THE PROBLEM	1
	Introduction	1
	Statement of the Problem	1
	Importance of the Study	1
	Overview of Procedures	3
	Definition of Terms	3
	Organization of Chapters	4
II.	REVIEW OF THE LITERATURE	5
	Philosophy of the Use of Teacher Ratings	5
	Student Academic Achievement and Teacher Ratings	7
	Summary of the Literature	10
III.	PROCEDURE	11
IV.	RESULTS	14
	Results of Analysis of Variance: Full Score	14
	Results of Analysis of Variance: Teacher Competency	19
	Results of Analysis of Variance: Teacher Personality	23
V.	SUMMARY AND CONCLUSIONS	28
	Summary and Conclusions	28
	Suggestions for Further Research	34
BIBLIOGRAPHY		37
APPENDIX		39

LIST OF TABLES

TABLE		PAGE
I.	Analysis of Variance Summary: Full Score	15
II.	Duncan's New Multiple Range Test Applied to Full Score Rating Mean Scores for Groups Formed by Class and Sex . .	17
III.	Duncan's New Multiple Range Test Applied to Full Score Rating Mean Scores for Groups Formed by Class and Discrepancy Score	18
IV.	Analysis of Variance Summary: Teacher Competency Scores . .	20
V.	Duncan's New Multiple Range Test Applied to Teacher Competency Rating Mean Scores for Groups Formed by Class and Sex	21
VI.	Duncan's New Multiple Range Test Applied to Teacher Competency Rating Mean Scores for Groups Formed by Class and Discrepancy Score	22
VII.	Analysis of Variance Summary: Teacher Personality Scores .	24
VIII.	Duncan's New Multiple Range Test Applied to Teacher Personality Rating Mean Scores for Groups Formed by Class and Sex	25
IX.	Duncan's New Multiple Range Test Applied to Teacher Personality Rating Mean Scores for Groups Formed by Class and Discrepancy Score	27

ABSTRACT

This investigation attempted to determine significant differences between teacher ratings of students with a high positive discrepancy between course grade and GPA and those of students with a high negative discrepancy between course grade and GPA.

A scale constructed to evaluate teacher competency and teacher personality was administered to 368 students in Psychology 201 and 172 students in Psychology 302. Groups were composed of the top and bottom 27% discrepancy scores for males and females in each course. Teacher ratings by each group on full score, teacher personality, and teacher competency were used as basic data in three-way analysis of variance procedures.

It was found that significantly higher teacher ratings were given by the following groups: (1) students with high positive discrepancy scores in comparison to students with high negative discrepancy scores, and (2) students in Psychology 302 in comparison to students in Psychology 201.

Significant interaction indicated that, in comparison to other groups formed by sex and class, females in Psychology 302 gave higher teacher ratings than other groups except Psychology 302 males, and Psychology 201 females rated teachers lower than other groups. In comparison to other groups formed by class and discrepancy score, the Psychology 201 negative discrepancy score group gave significantly lower teacher ratings, and the Psychology 302 negative discrepancy score group gave higher teacher competency ratings.

CHAPTER I

THE PROBLEM

Introduction

Although the rating of college instructors by their students is a controversial issue, the practice has gained support in the form of student attempts to become involved in the administrative procedures of universities. Rationale supporting teacher ratings by students maintains that students are the "consumers" of educational endeavors and are in the best position to be aware of day-by-day classroom occurrences. However, in spite of various attempts by researchers to investigate variables influencing the validity of ratings, such validity is still questionable and makes the use of teacher ratings debatable.

Statement of the Problem

It was the intent of this study to investigate one facet of the validity of student ratings of teachers. Since opponents of teacher ratings claim that students reveal personal bias or prejudice in their ratings, it was the purpose of this study to determine if there are significant differences in teacher ratings between students who receive a course grade higher than their GPA and students who receive a course grade lower than their GPA.

Importance of the Study

Although research has been directed toward several dimensions of teacher rating including the relationship of course grade to teacher

ratings, it is suggested that course grade-teacher rating research may yield different conclusions from research based on a difference between course grade and GPA. Research based on course grade alone does not reveal sufficient information concerning possible reasons for the student's perception of his instructor. One may speculate that a student may be satisfied with poor course grades if his general achievement level is low and thus rate the teacher no differently from students who consistently receive high course grades.

However, this research proposes to investigate teacher ratings by comparing teacher ratings of students who receive a psychology course grade higher than their achievement level with ratings of students who receive a psychology course grade lower than their achievement level.

Consideration of the foregoing led to the formulation of the following hypotheses, which have been stated in the null form:

1. There will be no significant differences in ratings of teacher personality, teacher competency, or full score between students with high positive discrepancy scores and students with high negative discrepancy scores.
2. There will be no significant differences in ratings of teacher personality, teacher competency, or full score between students enrolled in Psychology 201 and students enrolled in Psychology 302.
3. There will be no significant differences in ratings of teacher personality, teacher competency, or full score between male students and female students.
4. There will be no significant differences in ratings of teacher personality, teacher competency, or full score

when the following factors are considered simultaneously:

(a) sex and classes
(b) sex and discrepancy scores
(c) class and discrepancy scores
(d) sex, class, and discrepancy scores

Overview of Procedures

During the final week of winter quarter of the 1969-70 school year, teacher rating sheets were administered to five sections of Psychology 201 and four sections of Psychology 302.

The ratings were scored on the basis of an overall rating, as well as subratings of teacher competency and teacher personality. Discrepancy scores were computed by finding the difference between each student's course grade and his GPA. Positive and negative discrepancy score groups were identified, and subgroups were formed by finding the top and bottom 27% discrepancy scores for males and females within each course.

Subgroup full scores, teacher competency scores, and teacher personality scores were used as data upon which an analysis of variance was based to determine if significant differences existed among the student groups.

Definition of Terms

Positive discrepancy score. This term refers to the difference between a course grade and a GPA which is lower than the course grade.

Negative discrepancy score. This term refers to the difference between a course grade and a GPA which is higher than the course grade.

High positive discrepancy score. A high positive discrepancy score is a score in the top 27% of positive discrepancy scores.

High negative discrepancy score. A high negative discrepancy score is a score included in the lower 27% of negative discrepancy scores.

Psychology 201. This course number refers to the first course in general psychology. The majority of psychology 201 students are freshmen and sophomores.

Psychology 302. This course number refers to Educational Psychology, a required course in the curriculum for teacher preparation. The majority of students are juniors and seniors.

GPA. GPA is a term used to denote cumulative grade point average and, in this study, referred to the grade point average recorded at the end of winter quarter, 1969-70.

Organization of Chapters

The study is introduced in Chapter I, which contains an introduction; a statement of the problem; a statement of the hypotheses; an overview of procedures; the definition of terms; and a description of the organization of the chapters.

Chapter II is a review of literature pertinent to this study, including literature concerning the philosophy of the use of teacher ratings and student achievement-teacher rating research.

Chapter III is a description of the procedures used in the study, and Chapter IV is a presentation of the results of analysis of variance.

The final chapter, Chapter V, summarizes the study, states the conclusions reached as a result of the findings, and concludes with suggestions for further research.

CHAPTER II

REVIEW OF THE LITERATURE

The controversial nature of the evaluative use of teacher ratings by students has stimulated various authors to attempt the justification of their varied philosophies of teacher ratings through narrative efforts as well as through an experimental approach. Researchers have sought to clarify the questionable validity of ratings in the literature of the past thirty or forty years. For purposes of this study, the most pertinent facets concerning student ratings of teachers appeared to be: first, the general philosophy concerning the acceptance or rejection of the use of teacher ratings; and second, the relationship between student academic achievement and teacher ratings.

Philosophy of the Use of Teacher Ratings

Ryans' (1954) observations concerning teacher attitudes towards ratings revealed the following four reasons why teachers object to being assessed or judged by their students: (1) Teachers claim they are "professional persons" and therefore should not be subjected to performance evaluation; (2) Since many teachers are introverted and mildly insecure, they seek to avoid the potentially ego-damaging influence; (3) Raters may be prejudiced and employ ratings to reflect their bias or prejudice; and (4) Student ratings are subjective and often unreliable. Although Ryans excludes the first two reasons as invalid, he is in agreement with most critics of teacher ratings whose

opposition is based upon the last two premises. In spite of the fact that Ryans recognizes the weaknesses of ratings, he advocates their use in the belief that discriminating judgments of performance are necessary wherever individual differences in performance exist.

Eckert (1950) reviews various methods of evaluating college teaching including published materials, participation in professional organizations, and ratings by fellow workers. Included in her review is an examination of teacher ratings by students, and she suggests that the quality of learning may be improved if students are challenged to think about educational problems as they rate their teachers. The use of ratings is advocated as a more democratic administration of the college since students as well as faculty and administration would share in furnishing the evidence which influences promotions, salary increases, etc..

In an attempt to discover student concepts of good teaching, Coffman (1954) investigated values reflected in the overall ratings of approximately two thousand students. The results of factor analysis indicated that students value understanding of the learner, organization of the course, and verbal fluency on the part of their teachers. Coffman concluded that these highly valued factors are sound values, therefore making the use of ratings more acceptable.

Mueller (1951) investigated trends in student ratings of faculty in 269 institutions which reported experience with ratings and concluded that ratings were used extensively. He found that institutions with enrollments under 500 tended to be less satisfied with faculty rating experiences than did larger institutions. Mueller advocated the scholarly investigation of various facets of ratings so that their value may be ascertained.

Starrak (1934) analyzed the results of a teacher ratings scale used over a period of years at Iowa State College. Since a large percentage of the faculty was rated several times in successive quarters, it was found that ratings received by teachers tended to increase with successive ratings so that many teachers gradually moved from the lower quartile into the second and third quartiles over a two-year period. Although the validity of the scale was still in doubt, one may safely conclude that during the period studied, teachers tended to raise themselves in the opinion of the students.

In response to requests for information on student rating of faculty, McKeachie (1969) was commissioned to prepare an article for the Bulletin of the American Association of University Professors. In addition to presenting a sample evaluation form, McKeachie recognized the increasing demand for ways of evaluating teaching, which he attributed to an increased interest in college teaching. He stated that educational objectives are rarely defined clearly enough to permit measurement, and even when they are, we seldom have adequate measures. Summarizing research on student ratings, McKeachie concluded that they do have some validity because teachers rated as effective by students tend to be those whose students learn most as evidenced in subject matter attitude sophistication scores and performance on a test of psychological thinking.

Student Academic Achievement and Teacher Ratings

Carl Weaver (1960) examined ratings on both personality of the instructor and teaching skills and abilities of the instructor in an attempt to determine whether students tend to give the instructor about

the same kind of grade they expect to receive in the course. He concluded that the ratings of his 699 subjects were biased in the direction of the grades they designated as expecting to receive, and that most of the student bias was directed toward teaching skills rather than instructor personality.

Nichol's (1967) study of the various variables involved in student ratings of teacher effectiveness suggests that there is no consistent basis for expecting students with a 3.00 GPA or above to rate teachers significantly higher than students with below a 3.00 GPA. Neither was a consistent pattern found encompassing significant differences in ratings between students with a 3.50 GPA or above and those with a below a 2.00 GPA.

In an attempt to determine whether any significant relationship is present between student's known level of course achievement and ratings of their instructor, Bendig (1953) concluded the following: (1) Classes whose mean achievement is high show negative correlations between individual student ratings of the instructor and individual student ratings of the instructor and individual student achievement; (2) Classes whose mean achievement level is low show positive correlations between individual student ratings of teachers and individual student achievement. Bendig suggests that although individual student achievement may be ignored in evaluating ratings, the mean achievement of the class doing the rating must be considered.

Correlations of grades and student ratings for 409 students under eleven instructors showed that although the correlations ranged from -.860 to +.890, the average correlation was .070. Remmers (1930) concluded that for the average student and the average instructor, there

is practically no relationship between the student's grades and his judgment of the instructor.

Starrak (1934) found a correlation coefficient of .15 $\pm$ .031 between course grade and teacher ratings in his analysis of a teacher rating scale used at Iowa State College. He concluded that the ability of the student does not seem to appreciably influence the rating which he gives his instructor.

Anikeef (1953) ranked nineteen faculty members in accordance with the rating scores assigned to them by their students. Ratings were correlated with grading leniency and student absence extensiveness of the same instructors. Anikeef found that grading leniency correlated highest with freshman-sophomore rating scores and lowest with junior-senior rating scores. Student absence extensiveness correlated negatively on all academic levels, but the correlation was significant only on the combined four-year breakdown.

Russell and Bendig (1953) divided students into three groups based on a difference between their obtained course grade and a grade predicted by a regression equation based on A.C.E. scores. Members of the "plus" group had obtained grades more than one-half standard error of estimate above their predicted grade; the "equal" group had obtained grades within one-half standard error of estimate of their predicted grade; and members of the "minus" group had obtained grades more than one-half standard error of estimate below their expected grade. No overall difference was found between the "plus" and "minus" groups in their instructor ratings on the Miami Instructor Rating Sheet. However, the "plus" group gave more favorable ratings on scales measuring attitude toward the course.

Summary of the Literature

The use of student evaluation of teachers is generally supported by the literature. Authors base their support on premises and assumptions such as: (1) students are the consumers of educational endeavors and therefore are the "experts" in evaluating teacher effectiveness; (2) we are forced to rely upon evaluation of teachers by students since more valid measures of assessing educational goals and objectives are rarely available; (3) students should be allowed more involvement in the administrative decisions affecting teacher tenure, promotions, etc.; and (4) the use of teacher ratings may stimulate student thinking about the nature and purposes of higher education.

The majority of research investigating the relationship between academic achievement and teacher ratings indicates that there is little, if any, correlation between grades and teacher ratings. However, consideration of factors such as mean achievement level of class and size of college were suggested as factors which may influence rating of teachers.

CHAPTER III

PROCEDURE

A teacher rating scale constructed to include items directly related to the areas of teacher competency and teacher personality was administered during the final week of winter quarter 1969-70 to students in four sections of Psychology 302 and to students in five sections of Psychology 201. The scale, which may be found in Appendix A, provided for the rating of five aspects of teacher competency such as organization, quality of explanations and ability to stimulate curiosity and thought. The instructor's personality was rated by five items including such items as sensitivity to student problems, fairness, and tolerance of disagreement.

Three instructors, two of whom taught two sections of Psychology 201 and two instructors, each of whom taught two sections of Psychology 302 were rated. All instructors included in this study were male, either at or near the doctoral level; and the 302 and 201 instructors were equated as closely as possible on years of teaching experience.

Since the rating was done during the final week of the second quarter of the university year, the raters had had at least one quarter's experience with the instructor rated. It was assumed that every student had some expectation of the grade he would receive in the course as a result of mid-quarter examinations and other grades.

The researcher distributed the rating sheets to students, explained that the ratings would be used for research, and remained in the classroom while the ratings were being marked. None of the instructors was present while the students were completing the rating sheets; and since the ratings were signed, the students were assured that the results of their ratings would be available to the instructors only in the form of group data.

The ratings were scored by the researcher using a method which allowed scores of four, three, two, and one for the degrees of answers for each item. A maximum score of twenty was possible for both teacher competency and teacher personality, and a full score of 48 was possible. Teacher competency and teacher personality scores as well as a full scale score were recorded for each student.

When final course grades were available, the course grade for each student rater was obtained from his respective teacher's record book. Cumulative grade point averages including winter quarter 1969-70 were obtained from the Office of the Registrar, and discrepancy scores were computed and recorded on the rating sheets.

Four groups of students were identified in both Psychology 201 and Psychology 302 by separating the males and the females and by further dividing these two groups into positive discrepancy score groups and negative discrepancy score groups. The top 27% and bottom 27% discrepancy scores were identified in order to form the eight groups included in the research. These groups are and will be referred to as: Psychology 201 positive females, Psychology 201 negative females, Psychology 201 positive males, Psychology 201 negative males, Psychology

302 positive females, Psychology 302 negative females, Psychology 302 positive males, and Psychology 302 negative males.

Teacher competency, teacher personality, and full scores were used as basic data which were analyzed by three-way analysis of variance procedures to determine if significant differences existed among the groups.

CHAPTER IV

RESULTS

Rating sheets completed by 578 students were collected. Twenty nine ratings were discarded because the age of the rater exceeded the predetermined limit of twenty four years, as it was felt that their inclusion would make the sample less homogeneous. Nine papers were discarded either because the rater had entered a pseudonym or had failed to follow directions. Several raters wrote comments and explanations in the margin, suggesting their eagerness to communicate accurately and their sincerity in performing the task.

The rating sheets were sorted into male and female groups within each class and each group aligned by discrepancy scores from most positive to most negative. The top 27% discrepancy scores and bottom 27% discrepancy scores were identified within each group, and these composed the following subgroups: Psychology 201 female negative, Psychology 201 female positive, Psychology 201 male negative, Psychology 201 male positive, Psychology 302 female negative, Psychology 302 female positive, Psychology 302 male negative,and Psychology 302 male positive. The teacher ratings of each group furnished basic data for the three-way analysis of variance.

Results of Analysis of Variance: Full Score

Data in Table 1 show the analysis of variance summary for full score teacher ratings. The resulting F of 11.35 between classes was

TABLE 1
ANALYSIS OF VARIANCE SUMMARY:
FULL SCALE SCORES

Source	Sum of Squares	df	Mean Square	F
Sex	68.61	1	68.61	2.82
Classes	276.07	1	276.07	11.35****
Discrepancy Scores	249.80	1	249.80	10.27***
Sex x Discrepancy Scores	.70	1	.70	.03
Sex x Classes	180.20	1	180.20	4.45*
Class x Discrepancy Scores	128.48	1	128.48	5.28*
Sex x Class x Discrepancy Scores	18.37	1	18.37	.76
Error	6958.93	286	24.33	
Total	7809.16	293		

* denotes significant differences at the .05 level of confidence
*** denotes significant differences at the .005 level of confidence
**** denotes significant differences at the .001 level of confidence

significant at the .001 level of confidence. A mean rating of 40.21 by Psychology 302 students in comparison with a mean rating of 38.14 by Psychology 201 students indicated that Psychology 302 students tended to rate their teachers significantly higher than Psychology 201 students.

The F of 10.27 between the positive discrepancy score group and the negative discrepancy score group demonstrated significant differences at the .005 level of confidence on full score ratings. The positive discrepancy score group, with a mean rating of 39.72, rated teachers higher than the negative discrepancy score group which gave a mean rating of 37.88.

An F of 4.45, which is significant at the .05 level of confidence, resulted on sex x classes interaction. A Duncan's New Multiple Range Test was applied to the means of the teacher rating full scores to ascertain which group means differed significantly from one another. This test revealed that, at the .05 level of confidence, the mean rating of Psychology 201 females was significantly lower than the following groups: Psychology 201 males, Psychology 302 males, and Psychology 302 females. Table 2 contains these results.

An F of 5.28 on the class x discrepancy score interaction was significant at the .05 level of confidence. A Duncan's New Multiple Range Test, the results of which are presented in Table 3, was applied to the means of the positive discrepancy score and negative discrepancy score groups. This test revealed that although higher ratings were associated with positive discrepancy score groups, there were significant differences between the mean full score ratings of the Psychology 201 negative group and each of the other groups. Rating scores of the

TABLE 2
DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO
FULL SCORE RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND SEX

	Means	A 37.46	B 39.19	C 39.63	D 40.83
A	37.46		1.73*	2.17*	3.37*
B	39.19			.44	1.64
C	39.63				1.20

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 females
B	Psychology 201 males
C	Psychology 302 males
D	Psychology 302 females

TABLE 3
DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO
FULL SCORE RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND DISCREPANCY SCORE

	Means	A 36.76	B 39.51	C 40.17	D 40.26
A	36.76		2.75*	3.41*	3.50*
B	39.51			.66	.75
C	40.17				.09

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 negative
B	Psychology 201 positive
C	Psychology 302 positive
D	Psychology 302 negative

Psychology 201 negative group were significantly lower than those of the other groups.

Results of Analysis of Variance: Teacher Competency

The analysis of variance summary of teacher competency ratings is found in Table 4. The between classes F of 20.83 was significant at the .001 level of confidence. Psychology 302 classes, with a mean teacher competency rating of 16.39, rated teachers significantly higher than did the Psychology 201 classes with a mean teacher competency rating of 14.86.

Significant differences at the .05 level of confidence were revealed by the F of 4.81 which was found between the teacher competency ratings of the positive discrepancy score group and the negative discrepancy score group. An examination of the group means revealed a mean of 15.69 for the positive discrepancy score group and a mean of 15.01 for the negative discrepancy score group.

The F of 13.33 on sex x class interaction was found to be significant at the .001 level of confidence. A Duncan's New Multiple Range Test was performed on the means of the groups, the results of which are in Table 5. This test revealed that the mean teacher competency rating of Psychology 302 females was significantly higher than that of all other groups.

A significant F at the .005 level of confidence was found on the class x discrepancy score interaction. Application of a Duncan's New Multiple Range Test revealed significant differences, as shown in Table 6, between the Psychology 201 negative group and all other groups, the Psychology 201 negative having the lowest mean. Significant

TABLE 4
ANALYSIS OF VARIANCE SUMMARY:
TEACHER COMPETENCY SCORES

Source	Sum of Squares	df	Mean Square	F
Sex	6.04	1	6.04	.84
Classes	150.39	1	150.39	20.83****
Discrepancy scores	34.70	1	34.70	4.81*
Sex x Classes	10.66	1	96.23	13.33****
Sex x Discrepancy Scores	96.23	1	10.66	1.48
Class x Discrepancy Scores	75.09	1	75.09	10.40***
Sex x Class x Discrepancy Scores	12.76	1	12.76	1.77
Error	2065.04	286	7.22	
Total	2450.91	293		

* denotes significant differences at the .05 level of confidence
*** denotes significant differences at the .005 level of confidence
**** denotes significant differences at the .001 level of confidence

TABLE 5
DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO
TEACHER COMPETENCY RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND SEX

	Means	A 14.48	B 15.45	C 15.63	D 17.20
A	14.48		.97*	1.15*	2.72*
B	15.45			.18	1.75*
C	15.63				1.57*

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 females
B	Psychology 201 males
C	Psychology 302 males
D	Psychology 302 females

TABLE 6
DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO
TEACHER COMPETENCY RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND DISCREPANCY SCORE

	Means	A 14.17	B 15.55	C 16.00	D 16.79
A	14.17		1.38*	1.83*	2.62*
B	15.55			.45	1.24*
C	16.00				.79

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 negative
B	Psychology 201 positive
C	Psychology 302 positive
D	Psychology 302 negative

differences between the Psychology 201 positive and Psychology 302 negative means were also found, the Psychology 302 negative mean being the highest of means for all groups.

Results of Analysis of Variance: Teacher Personality

The analysis of variance summary for teacher personality is found is Table 7. Analysis of variance procedures with regard to teacher personality ratings by classes revealed an F of 9.43 which was significant at the .005 level of confidence. Psychology 302 students gave a mean rating of 17.63 on teacher personality, and Psychology 201 students gave a lower mean score of 16.85 on teacher personality.

Significant differences between discrepancy score groups were found by an F of 6.77, which was significant at the .01 level of confidence. The positive discrepancy score group gave instructors a mean rating of 17.41 on teacher personality, and the negative discrepancy score group gave teachers a lower mean teacher personality rating of 16.79.

The F of 4.60 on sex x class interaction was significant at the .05 level of confidence. The Duncan's New Multiple Range Test, the results of which are shown in Table 8, revealed significant differences between Psychology 201 females and Psychology 302 females, as well as between Psychology 201 males and Psychology 302 females. An examination of the mean scores of each group showed that Psychology 302 females gave their teachers the highest mean rating of all groups: it was eighteen, only two points below a possible twenty point score on teacher personality.

TABLE 7
ANALYSIS OF VARIANCE SUMMARY:
TEACHER PERSONALITY SCORES

Source	Sum of Squares	df	Mean Square	F
Sex	1.31	1	1.31	.31
Classes	39.24	1	39.24	9.43***
Discrepancy Scores	28.16	1	28.16	6.77**
Sex x Discrepancy Scores	.89	1	.89	.21
Sex x Classes	19.14	1	19.14	4.60*
Class x Discrepancy	23.90	1	23.90	5.75*
Sex x Class x Discrepancy Scores	-.53	1	-.53	
Error	1189.62	286	4.16	
Total	1301.73	293		

* denotes significant differences at the .05 level of confidence
** denotes significant differences at the .01 level of confidence
*** denotes significant differences at the .005 level of confidence

TABLE 8
DUNCAN'S NEW MULTIPLE RANGE TEST APPLIED TO
TEACHER PERSONALITY RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND SEX

	Means	A 16.69	B 17.12	C 17.29	D 18.00
A	16.69		.43	.60	1.31*
B	17.12			.17	.88*
C	17.29				.71

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 females
B	Psychology 201 males
C	Psychology 302 males
D	Psychology 302 females

A significant F at the .05 level of confidence was found in class x discrepancy score interaction. Application of the Duncan's New Multiple Range Test showed significant differences between the Psychology 201 negative discrepancy score group and all other groups. Table 9 indicates that the means of Psychology 201 positive, Psychology 302 positive, and Psychology 302 negative groups were respectively 17.36, 17.53, and 17.74 in comparison to 16.35 for the Psychology 201 negative group.

TABLE 9
DUNCAN'S NEW MULTIPLE RANGE REST APPLIED TO
TEACHER PERSONALITY RATING MEAN SCORES FOR GROUPS
FORMED BY CLASS AND DISCREPANCY SCORE

	Means	A 16.35	B 17.36	C 17.53	D 17.74
A	16.35		1.01*	1.18*	1.39*
B	17.36			.17	.38
C	17.53				.21

* denotes significant differences at the .05 level of confidence

Key	Group
A	Psychology 201 negative
B	Psychology 201 positive
C	Psychology 302 positive
D	Psychology 302 negative

CHAPTER V

SUMMARY AND CONCLUSIONS

This investigation attempted to determine whether there are significant differences in the full score, teacher competency, and teacher personality ratings between students with high positive discrepancy scores and students with high negative discrepancy scores.

The sample was composed of 540 students who were enrolled in either Psychology 201 or Psychology 302 at Appalachian State University. A teacher rating scale constructed to include items directly related to teacher competency and teacher personality was administered during the final week of winter quarter of 1969-1970 to students enrolled in five sections of Psychology 201 and four sections of Psychology 302.

Teacher rating sheets were scored and discrepancy scores computed by finding the difference between the student's course grade and his overall GPA. Sample subgroups were formed by finding the top and bottom 27% of discrepancy scores for males and females in Psychology 201 and for males and females in Psychology 302.

The three variables of sex, class, and discrepancy score were analyzed by a 2x2x2 factorial design. These analyses were made to test the following hypotheses:

1. There will be no significant differences in ratings of teacher personality, teacher competency or full score between students with high positive discrepancy scores and students with high negative discrepancy scores.

2. There will be no significant differences in ratings of teacher personality, teacher competency or full score between students enrolled in Psychology 201 and students enrolled in Psychology 302.
3. There will be no significant differences in ratings of teacher personality, teacher competency or full score between male students and female students.
4. There will be no significant differences in ratings of teacher personality, teacher competency, or full score when the following factors are considered simultaneously:

    (a) sex and classes
    (b) sex and discrepancy scores
    (c) class and discrepancy scores
    (d) sex, class, and discrepancy scores

As a result of analysis of variance, the first hypothesis was rejected since significant differences in rating scores were found between the positive discrepancy score group and the negative discrepancy score group. Students with high positive discrepancy scores tended to rate teachers higher than students with high negative discrepancy scores on full score, as well as teacher personality and teacher competency. The greatest difference in ratings of the two groups was at the .005 level of confidence on full score ratings, and the least difference was at the .05 level of confidence on teacher competency. The positive discrepancy score group had a slightly smaller standard deviation than the standard deviation of the negative discrepancy score group, indicating greater agreement among students who made higher course grades than their GPA.

The second hypothesis was rejected since analysis of variance procedures revealed significant differences between the ratings of Psychology 201 students and Psychology 302 students. Higher ratings in all three teacher rating categories were associated with Psychology 302 students. Differences between the ratings of Psychology 201 and Psychology 302 students were significant at the .001 level of confidence for teacher competency and full score, and teacher personality rating differences were significant at the .005 level of confidence. An examination of the group standard deviations revealed no pattern which indicated greater agreement among one group than the other.

It is suggested that the difference in average class size for each of the courses may have influenced differential rating between classes. Each section of Psychology 201 had approximately ninety five students enrolled, whereas the average number of students in each section of Psychology 302 was approximately fifty. Perhaps lower teacher ratings by Psychology 201 students could be attributed to the impersonality generally associated with larger class size.

The third hypothesis was accepted since no significant differences were found between the ratings of males and females.

Since significant differences were found in sex x class interaction and in class x discrepancy score interaction, the fourth hypothesis was partially rejected. A review of sex and class interaction revealed that Psychology 201 females rated teachers significantly lower than other groups on full score and teacher competency, but teacher personality ratings of Psychology 201 females were significantly lower than only one group--the Psychology 302 females. We might conclude that low teacher personality ratings are less associated with

Psychology 201 females than are low full score and teacher competency ratings.

Sex and class interaction analysis also revealed significant differences between the mean score of Psychology 302 females and all other groups on teacher competency, as well as between mean scores of Psychology 302 females and both Psychology 201 groups on teacher personality. In each case, ratings by Psychology 302 females were higher. Although Psychology 302 females had a significantly higher mean score on the full rating scale than Psychology 201 females, there was less difference between Psychology 302 females and other groups on this score than on teacher personality or teacher competency scores.

The nature of the course content of each of the courses involved might possibly have influenced the ratings given by females. While there were no significant differences in ratings between the male groups, female Psychology 201 groups were associated with low ratings; and female Psychology 302 groups were associated with high ratings. Perhaps the humanistic, career-oriented nature of Psychology 302 is more meaningful to females than the detached, empirical material presented in Psychology 201. These course content differences may be related to teacher ratings in the respective courses.

Class and discrepancy score interaction revealed significantly lower ratings of full score, teacher personality, and teacher competency by the Psychology 201 negative discrepancy score group. In all three teacher rating categories, Psychology 201 negative mean scores were significantly lower than mean score of other groups. The only additional significant difference was found between the Psychology 201 positive group and the Psychology 302 negative group, with the

Psychology 302 negative group having the higher mean score, so that the Psychology 302 negative group had a higher mean score than either the Psychology 201 positive or Psychology 201 negative groups on teacher competency.

It may be concluded that the Psychology 201 negative discrepancy score group may be associated with lower teacher ratings on full score, teacher competency, and teacher personality. The association of low ratings with the Psychology 201 negative group might be attributed to the disparity between the students' course grades and their overall higher academic achievement level. Perhaps these negative discrepancy score students attributed their lack of success in the course to the teacher.

In contrast to the association between Psychology 201 negative discrepancy score students and lower teacher ratings, the Psychology 302 negative discrepancy score group may be associated with significantly higher teacher competency ratings than either of the Psychology 201 groups, but not significantly higher than the Psychology 302 positive group. One might speculate that those Psychology 302 students who received course grades lower than their GPA might have experienced an attitude change toward the nature of the course and thus developed an appreciation of the instructor's competency as reflected in high teacher competency ratings.

The following conclusions were drawn from these data:

1. In this sample, students with high positive discrepancy scores tended to rate teachers higher on full score, teacher personality, and teacher competency than students with high negative discrepancy scores.

2. Students in Psychology 302 tended to rate teachers higher on full score, teacher personality, and teacher competency than students in Psychology 201.

3. Psychology 201 females tended to give lower ratings on full score and teacher competency than other groups formed by class and sex, and lower teacher personality ratings than Psychology 302 females.

4. Psychology 302 females tended to give higher teacher competency ratings than all other groups formed by sex and class and higher teacher personality ratings than the Psychology 201 males and females.

5. The Psychology 201 negative discrepancy score group tended to rate teachers lower on full score, teacher personality, and teacher competency than other groups formed by class and discrepancy score.

6. The Psychology 302 negative discrepancy score group tended to give higher teacher competency ratings than the Psychology 201 discrepancy score groups.

The results of this study support the conclusion that teacher rating scores may be partially a function of sex, course, and a discrepancy between the course grade and GPA. The combination and interaction of these variables resulted in significant differences in teacher ratings in this study. One might note that, although there were some isolated instances of variables affecting only one of the three teacher rating categories--teacher personality, teacher competency, and full score, the general trend was for all three ratings categories to be affected by one of the independent variables. For example, negative discrepancy scores were associated with not only low ratings for full score, but also low ratings for teacher personality and teacher competency.

The results of this study indicate that general statements concerning the validity of student ratings may be meaningless unless the specific variables affecting the ratings are identified. Ratings should be evaluated in light of the course content, the particular student population composition, mean achievement level of the class, and individual student achievement as related to his course grade.

Suggestions for Further Research

This author is in complete agreement with Muller (1951) who advocates further investigation of the various facets of teacher ratings so that their value may be ascertained.

First, it is suggested that more information is needed concerning ratings between classes, between courses, and between departments. A wide range of courses should be included within each class. Rather than use only Psychology 302 to represent upper level students, other junior and senior courses should be included to determine if significant differences exist among freshman-sophomore and junior-senior groups.

One may speculate that there may be significant differences in teacher ratings from course to course. By determining the presence or lack of differences in teacher ratings between courses, ratings could be interpreted more meaningfully. An expedient method of studying between-course differences might utilize a set of ratings of teachers made when they teach each course in a series such as the Psychology 301, 302, 303 sequence at Appalachian State University.

Teacher rating research has been largely limited to ratings of psychology instructors. It is suggested that information is needed

concerning teacher rating in other university departments and that inter-departmental analysis should be done.

Second, the relationship between course grade, general academic level and teacher ratings needs further investigation. In addition to utilizing high positive and high negative discrepancy score groups, an intermediate group might be formed for analysis.

Third, in order to validate the assumption that each student is aware of his course grade, teacher ratings might be collected after final course grades are posted. The effect of course grade on teacher ratings might also be determined by obtaining ratings both before and after final course grades are available and making a comparison of each student's two ratings in relation to his course grade.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anikeef, Alexis M. "Factors Affecting Student Evaluation of College Faculty Members," Journal of Applied Psychology, XXXVII (1953), 458-460.

Bendig, A. W. "The Relation of Level of Course Achievement to Students' Instructor and Course Ratings in Introductory Psychology," Educational and Psychological Measurement, XIII (1953), 437-447.

Coffman, W. E. "Determining Student's Concepts of Effective Teaching from their Ratings of Instructors," Journal of Educational Psychology, XXXXV (1954), 277.

Eckert, Ruth E. "Ways of Evaluating College Teaching," School and Society, LXXI (1950), 65-69.

McKeachie, W. J. "Student Ratings of Faculty," Bulletin of the American Association of University Professors, LV (December, 1969), 439-444.

Muller, F. J. "Trends in Student Ratings of Faculty," Bulletin of the American Association of University Professors, XXXVII (1951), 319-324.

Nichols, M. G. "A Study of the Influence of Selected Variables Involved in Student Evaluations of Teacher Effectiveness," Dissertation Abstracts, (February 1969), 2908-A.

Remmers, H. H. "To What Extent Do Grades Influence Student Ratings of Instructors?" Journal of Educational Research, XXI (1930), 314-316.

Russell, H. E. and A. W. Bendig. "An Investigation of the Relation of Student Ratings of Psychology Instructors to their Course Achievement When Academic Aptitude is Controlled," Educational and Psychological Measurement, XIII (1953), 626-635.

Ryans, D. G. "Notes on the Rating of Teacher Performance," Journal of Educational Research, XXXXVII (1954), 695-703.

Starrak, J. A. "Student Rating of Instructors," Journal of Higher Education, V (1934), 88-90.

Weaver, Carl H. "Instructor Rating by College Students," Journal of Educational Psychology, LI (1960), 21-25.

APPENDIX A

FACULTY EVALUATION

Name: _____

Age: _____

Sex: _____

Class: _____

Place an "X" in ONE of the blanks in each item.

1. Does he present the material in a well-organized fashion?
   _____well-organized
   _____adequate, could be better
   _____inadequate organization
   _____confused and unsystematic

2. Does he have a thorough knowledge and understanding of his subject matter?
   _____exceedingly well informed
   _____adequately informed
   _____not well informed
   _____very inadequately informed

3. Are his explanations clear and understandable?
   _____explanations clear and to the point
   _____explanations usually adequate
   _____explanations often inadequate
   _____explanations seldom given or usually inadequate

4. Does he stimulate curiosity?
   _____inspires students to independent thought
   _____occasionally inspiring; creates mild interest
   _____rarely inspiring
   _____destroys interest in subject

5. What is his attitude toward the subject?
   _____enthusiastic, enjoys teaching
   _____rather interested
   _____rather bored, routine interest
   _____not interested

6. Are the tests fair?
   _____very fair
   _____sometimes fair
   _____rarely fair
   _____very unfair

7. Is he fair in grading procedures?
   _____always fair in grading
   _____usually fair
   _____occasionally unfair
   _____unfair

8. Does he make students feel free to ask questions, disagree, express their own ideas, etc.?
   _____encourages student ideas
   _____tolerant of student ideas
   _____discourages student ideas
   _____intolerant, allows no contradiction

9. Is he approachable outside of class?
   _____consultations outside class welcomed
   _____consultations outside class tolerated
   _____consultations outside class discouraged
   _____consultations outside class avoided

10. Is he fair and impartial in treatment of all students?
   _____always fair and impartial
   _____usually fair and impartial
   _____often unfair and partial
   _____very unfair and partial

11. Does he appear sensitive to student's feelings and problems?
   _____responsive, actively concerned
   _____moderately sympathetic
   _____routine in attitude
   _____unaware, aloof, cold

12. How would you describe his personality?
   _____attractive personality, would like to know him better
   _____satisfactory personality
   _____rather unattractive personality
   _____extremely unattractive personality

What is your general estimate of this teacher?
   _____superior
   _____good
   _____average
   _____poor